Identifying Latent Concepts and Structures for Generalized Category Discovery

Boyang Dai, Chaoqi Chen, Yizhou Yu

香 港 大 學
THE UNIVERSITY OF HONG KONG

Forty-Third International Conference on Machine Learning

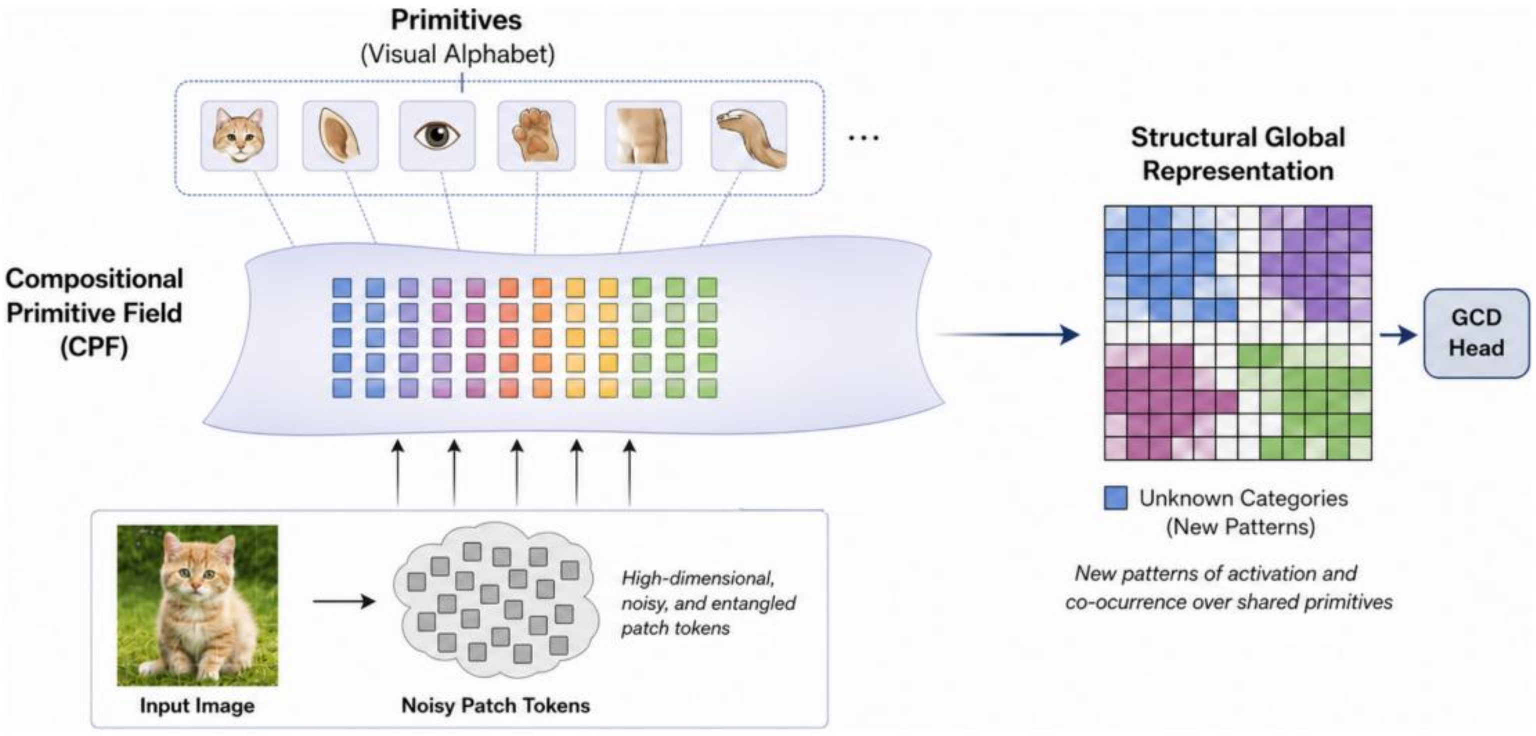
Motivation

- Raw patch tokens are high-rank, noisy, and entangled, making novel categories difficult to discover.

- CPF rewrites them through a compact set of reusable visual primitives:

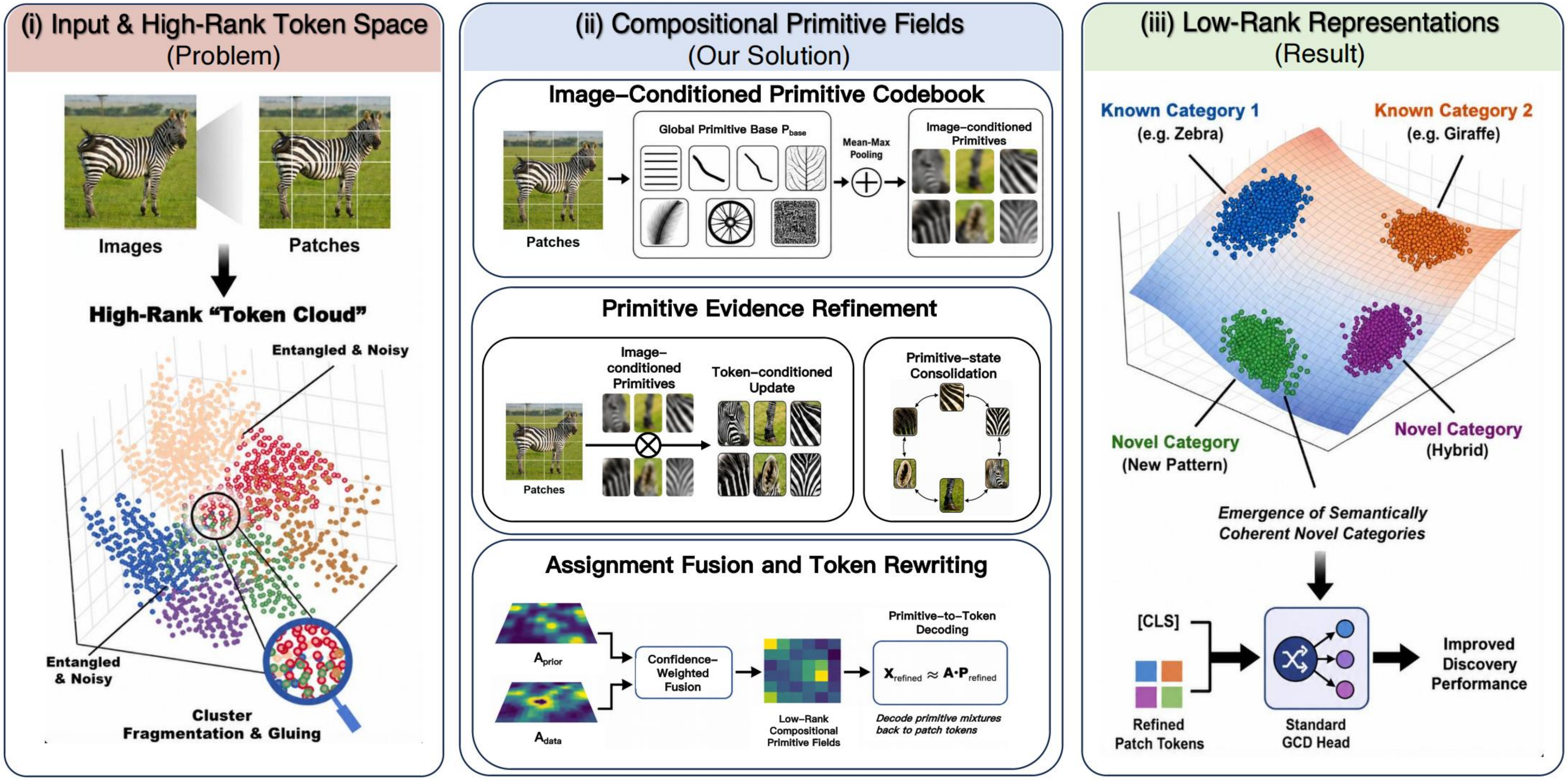
$$X \approx AP, \quad M \ll \min\{N, D\},$$

$$P \in \mathbb{R}^{M \times D}, \; A \in \mathbb{R}^{N \times M}$$



- P is a compact primitive codebook, and A encodes patch-to-primitive activations. Known and novel categories emerge as different combinations of the same primitive vocabulary.

Overview


(i) Input & High-Rank Token Space
(Problem)
Images
Patches
High-Rank "Token Cloud"
Entangled & Noisy
Entangled & Noisy
Cluster Fragmentation & Gluing
(ii) Compositional Primitive Fields
(Our Solution)
Image–Conditioned Primitive Codebook
Patches
Global Primitive Base P_base
Mean-Max Pooling
Image–conditioned Primitives
Primitive Evidence Refinement
Patches
Image–conditioned Primitives
Token–conditioned Update
Primitive–state Consolidation
Assignment Fusion and Token Rewriting
A_prior
A_data
Confidence–Weighted Fusion
Low–Rank Compositional Primitive Fields
Primitive–to–Token Decoding
X_refined ≈ A·P_refined
Decode primitive mixtures back to patch tokens
(iii) Low-Rank Representations
(Result)
Known Category 1
(e.g. Zebra)
Known Category 2
(e.g. Giraffe)
Novel Category
(New Pattern)
Novel Category
(Hybrid)
Emergence of Semantically Coherent Novel Categories
[CLS]
Refined Patch Tokens
Standard GCD Head
Improved Discovery Performance

Method

Codebook Construction

- Image-conditioned adaptation

$$\Delta P = \Phi(\mathrm{MeanMax}(X)),$$

$$P = P_{\mathrm{base}} + \Delta P$$

- Prior Assignment

$$\hat{X} = XW, \quad S = \frac{\hat{X}P^{\top}}{\sqrt{D}}$$

$$A_{\mathrm{prior}}(i,m) = \frac{\exp(S_{i,m})}{\sum_{j=1}^{N}\exp(S_{j,m})}$$

Primitive Evidence Refinement

- Token evidence pooling $P^0 = A_{\mathrm{prior}}^{\top}X$
- Token-conditioned update

$$Q_{TP} = P^0 W_q, \quad K_{TP} = XW_k, \quad V_{TP} = XW_v$$

$$P^1 = P^0 + \gamma W_{TP} V_{TP}$$

- Primitive-state consolidation

$$Q_{\mathrm{PP}} = \mathbf{P}^1 W_q', \quad K_{\mathrm{PP}} = \mathbf{P}^1 W_k', \quad V_{\mathrm{PP}} = \mathbf{P}^1 W_v'$$

$$P^2 = W_{PP} V_{PP}$$

Method

Assignment Fusion&Token Rewriting

- Data Assignment

$$\mathbf{A}_{\text{data}}^{(i,m)} = \text{Softmax}_m\left(\frac{1}{H}\sum_{h=1}^{H} W_{\text{TP}}^{(h,m,i)}\right)$$

- Confidence-weighted Fusion

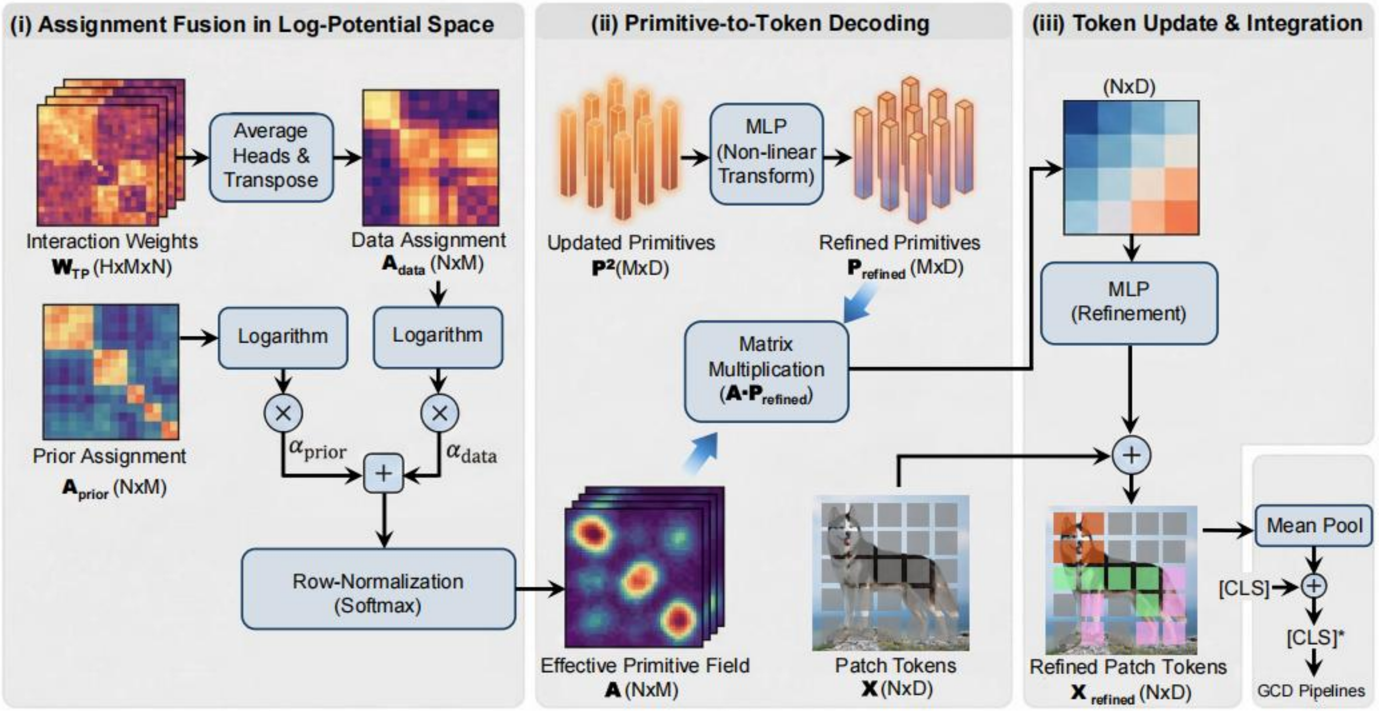
$$\mathbf{Z} = \alpha_{\text{prior}} \log \mathbf{A}_{\text{prior}} + \alpha_{\text{data}} \log \mathbf{A}_{\text{data}}$$

- Primitive-to-Token Decoding

$$\mathbf{X}_{\text{refined}} = \mathbf{X} + \text{MLP}(\mathbf{A} \cdot \mathbf{P}_{\text{refined}})$$

- Integration into GCD pipelines

$$[\mathbf{CLS}]^* = [\mathbf{CLS}] + \text{Mean}(\mathbf{X}_{\text{refined}})$$



Main Result

➢ **Fine-grained Datasets:**

Method	CUB-200			FGVC-Aircraft			Stanford-Cars			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	51.3	56.6	48.7	45.0	41.1	46.9	39.0	57.6	29.9	45.1	51.8	41.8
PromptCAL	62.9	64.4	62.1	52.2	52.2	52.3	50.2	70.1	40.6	55.1	62.2	51.7
AMEND	64.9	75.6	59.6	52.8	61.8	48.3	56.4	73.3	48.2	58.0	70.2	52.0
μGCD	65.7	68.0	64.6	53.8	55.4	53.0	56.5	68.1	50.9	58.7	63.8	56.2
ProtoGCD	63.2	68.5	60.5	56.8	62.5	53.9	53.8	73.7	44.2	57.9	68.2	52.9
InfoSieve	69.4	77.9	65.2	56.3	63.7	52.5	55.7	74.8	46.4	60.5	72.1	54.7
APL	64.5	68.1	62.1	56.6	60.2	54.8	**60.1**	77.6	51.2	60.4	68.6	56.0
ConGCD	68.1	68.5	67.8	59.7	61.3	**59.2**	59.1	**79.0**	49.8	62.3	69.6	58.9
SimGCD	61.2	65.8	58.9	54.5	59.3	52.1	54.6	72.8	45.7	56.8	66.0	52.2
+ CPF-GCD	63.9	65.5	63.1	55.8	59.1	54.2	57.6	73.7	49.8	59.1	66.1	55.7
	2.7	-0.3	4.2	1.3	-0.2	2.1	3.0	0.9	4.1	2.3	0.1	3.5
LegoGCD	61.9	71.9	56.9	54.6	62.7	50.6	53.7	72.2	44.9	56.7	68.9	50.8
+ CPF-GCD	66.0	73.5	62.0	56.3	62.3	53.2	59.7	73.6	**53.0**	60.7	69.8	56.1
	4.1	1.6	5.1	1.7	-0.4	2.6	6.0	1.4	8.1	4.0	0.9	5.3
CMS	65.7	75.8	60.7	50.8	62.1	45.1	51.3	73.5	40.6	55.9	70.5	48.8
+ CPF-GCD	66.8	74.8	62.9	51.5	60.4	47.0	57.9	78.2	48.0	58.7	71.1	52.6
	1.1	-1.0	2.2	0.7	-1.7	1.9	6.6	4.7	7.4	2.8	0.6	3.8
SelEx	75.6	77.3	74.7	61.1	68.7	57.3	55.5	77.6	44.8	64.1	74.5	58.9
+ CPF-GCD	**79.8**	**80.5**	**79.4**	**61.8**	**69.0**	58.1	58.9	**79.0**	49.2	**66.8**	**76.2**	**62.2**
	4.2	3.2	4.7	0.7	0.3	0.8	3.4	1.4	4.4	2.7	1.7	3.3
Avg. △	3.03	0.88	4.05	1.10	-0.50	1.85	4.75	2.10	6.00	2.95	0.83	3.98

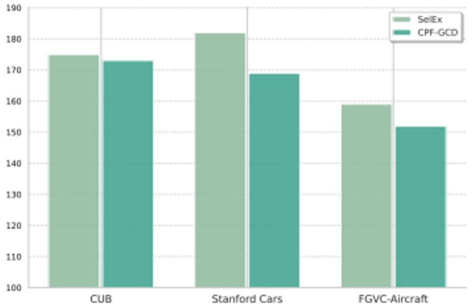
Main Result

- **Coarse-grained Datasets:**

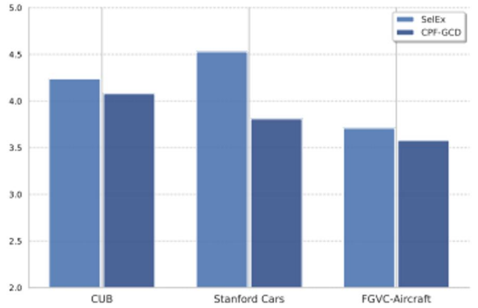
Method	CIFAR-10			CIFAR-100			ImageNet-100			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	91.5	**97.9**	88.2	73.0	76.2	66.5	74.1	89.8	66.3	79.5	88.0	73.7
PIM	94.7	97.4	93.3	78.3	84.2	66.5	83.1	95.3	77.0	85.4	92.3	78.9
PromptCAL	**97.9**	96.6	**98.5**	81.2	84.2	75.3	83.1	92.7	78.3	87.4	91.2	84.0
DCCL	96.3	96.5	96.9	75.3	76.8	70.2	80.5	90.5	76.2	84.0	87.9	81.1
ProtoGCD	97.3	95.3	98.2	81.9	82.9	80.0	84.0	92.2	79.9	87.7	90.1	86.0
InfoSieve	94.8	97.7	93.4	78.3	82.2	70.5	80.5	93.8	73.8	84.5	91.2	79.2
APL	97.1	94.9	98.2	80.9	81.6	79.5	83.2	92.6	78.5	87.1	89.7	85.4
ConGCD	97.4	95.2	**98.5**	82.5	85.9	77.3	85.9	93.4	82.5	88.6	91.5	86.1
SimGCD	96.5	96.1	96.7	80.1	82.1	76.0	83.2	94.0	77.8	86.6	90.7	83.5
+ CPF-GCD	97.4	95.5	98.3	80.9	82.0	78.8	84.2	92.7	80.0	87.5	90.1	85.7
	0.9	-0.6	**1.6**	**0.8**	-0.1	**2.8**	**1.0**	-1.3	**2.2**	**0.9**	-0.6	**2.2**
LegoGCD	96.8	96.1	97.2	81.9	83.5	78.7	**86.3**	94.5	82.1	88.3	91.4	86.0
+ CPF-GCD	97.5	95.8	98.3	82.8	83.3	**81.9**	86.1	93.1	**82.6**	**88.8**	90.7	**87.6**
	0.7	-0.3	**1.1**	**0.9**	-0.2	**3.2**	-0.2	-1.4	**0.5**	**0.5**	-0.7	**1.6**
CMS	95.2	96.9	94.4	82.4	**86.0**	75.3	83.1	94.1	77.6	86.9	92.3	82.4
+ CPF-GCD	96.2	97.0	95.7	**83.2**	85.8	78.1	85.1	**95.4**	79.9	88.2	**92.7**	84.6
	1.0	**0.1**	**1.3**	**0.8**	-0.2	**2.8**	**2.0**	**1.3**	**2.3**	**1.3**	**0.4**	**2.2**
SelEx	95.5	97.1	94.6	82.1	85.1	76.2	83.3	94.4	77.7	87.0	92.2	82.8
+ CPF-GCD	96.8	97.7	96.3	82.0	84.4	77.3	85.5	94.7	80.8	88.1	92.3	84.8
	1.3	**0.6**	**1.7**	-0.1	-0.7	**1.1**	**2.2**	**0.3**	**3.1**	**1.1**	**0.1**	**2.0**
Avg. △	**0.98**	-0.05	**1.43**	**0.60**	-0.30	**2.48**	**1.25**	-0.28	**2.03**	**0.95**	-0.20	**2.00**

Analysis

➢ Rank & Entropy Analysis

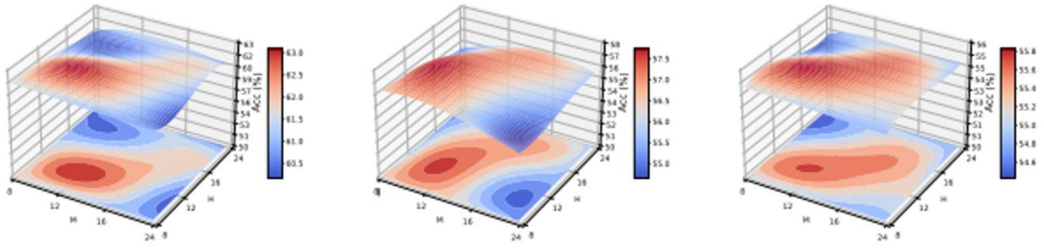


(a) $\mathrm{rank}_{99}(\mathcal{R})$ of fine-grained datasets



(b) $H(\mathcal{R})$ of fine-grained datasets

➢ Hyperparameter Sensitivity



(a) CUB *(b)* Stanford Cars *(c)* FGVC Aircraft

➢ Attention Visualization

➢ Primitive Visualization

Conclusion

- CPF identifies high-rank and entangled token geometry as a key bottleneck in Generalized Category Discovery.

- By rewriting patch tokens through low-rank compositional primitive fields, CPF produces compact and structured representations.

- As a plug-and-play module, CPF consistently improves known and novel category discovery across diverse GCD baselines, with extensive experiments, ablations, and visualizations verifying its effectiveness and generality.

Thank You !

More details are available in our paper and code.